

Received 11 May 2023, accepted 11 June 2023, date of publication 16 June 2023, date of current version 21 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3286937

RESEARCH ARTICLE

3D Reconstruction of Gastrointestinal Regions Using Single-View Methods

BILAL AHMAD¹, PÅL ANDERS FLOOR¹, IVAR FARUP¹, AND ØISTEIN HOVDE^{1,2,3}

¹Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

²Department of Gastroenterology, Innlandet Hospital Trust, 2819 Gjøvik, Norway

³Institute of Clinical Medicine, University of Oslo, 0372 Oslo, Norway

Corresponding author: Bilal Ahmad (bilal.ahmad@ntnu.no)

This work was supported by the Research Council of Norway through the Project CAPSULE under Grant 300031.

ABSTRACT Capsule endoscopy is about to become an alternative to traditional colonoscopy. One uses a wireless camera to visualize the gastrointestinal (GI) tract. A 3D model based on image sequences obtained from wireless capsule endoscopy (WCE) can be helpful to diagnose or analyse areas of interests. We have therefore investigated the possibility to provide enhanced viewing for gastroenterologists by reconstructing 3D shapes from WCE images. The study is done on virtual graphics-based models of human GI regions. The shape from shading (SFS) method is applied to colon images and the quality of the reconstructed shapes is compared with ground truth models. WCE images suffer from uneven and dim illumination due to point light source. Therefore, we provide a method based on surface normals from reconstructed 3D models to enhance contrast particularly in images capturing larger depths by changing the illumination from point light to directional light. Images of different resolution are also tested to evaluate their effect on the quality of the 3D reconstruction. We have also tested the shape from focus (SFF) method, a possibility for future WCEs, and compared the results with SFS. Finally, enhanced images and 3D shapes recovered with both methods have been evaluated by gastroenterologists through subjective experiments. Objective experiments indicate that both methods are capable of reconstructing the 3D shapes of colon images successfully, but the SFF method is better at retaining details in the reconstructed models than the SFS method. Subjective experiments show that contrast enhanced images are highly preferred over original images. Also, having the reconstructed 3D models in addition to the images during evaluation is found to be very useful by gastroenterologists and sometimes even being preferred over the original image.

INDEX TERMS 3D reconstruction, shape from shading, shape from focus, capsule endoscopy.

I. INTRODUCTION

It can be challenging to diagnose lower gastrointestinal (GI) diseases such as sources of gastrointestinal bleeding, large bowel cancer, ulcerative colitis and other disorders due to difficulty in accessing parts of the human GI system. Wireless capsule endoscopy (WCE) was introduced in year 2000 by Given Imaging [1] as an alternative to the regular colonoscopy. It is a patient-friendly, non-invasive, and a painless procedure to examine GI regions. There are several different types of WCEs available in the market that include esophageal, small bowel and colon capsules. WCE is a pill-sized capsule that the patient swallows. It has a camera

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan¹.

on board, capturing images of the GI region and transmitting them to a receiver. These images can be downloaded and reviewed by gastroenterologists to evaluate various diseases in patients.

A gastroenterologist usually spends hours on reviewing the WCE video. It is therefore necessary to provide a better perception of the diseased area in WCE images evaluation. A 3D model based on WCE frames can be helpful as it enables gastroenterologists to view the damaged area from different angles. A 3D model can also be helpful to better diagnose or analyse areas of interests, which may also help the patient to better understand the planned procedure.

3D reconstruction is a common problem in the field of computer vision that can be solved by applying different techniques to the images [2]. Vision-based depth estimation

methods are generally classified into different categories. Various monocular image-based techniques such as texture gradient analysis and photometric methods are used [3]. Other methods use multiple images and rely on the motion or multiple relative positions of the camera [4]. The use of 3D reconstruction is ubiquitous across many fields, such as in measurement systems, robotics, medical applications including diagnostics, video surveillance and monitoring etc. [5], [6].

Shape from shading (SFS) is a single-view method in computer vision that reconstructs the 3D shape of a scene or an object from one image. Therefore, it is useful in real-world applications where only one shot of the scene is available. WCE is one of the recent and useful applications of the SFS method, especially in areas where rapid movement of the capsule makes it difficult to capture images more than once.

The SFS method was first discussed by Horn and Brooks [7], who developed an iterative scheme based on nonlinear first-order PDE by relating 3D shape to intensity variations in one image. Ikeuchi and Horn employed stereographic projection rather than the gradient space to express surface orientation [8]. They introduced a brightness constraint to minimize the difference between the input image and the image produced by the reconstructed surface, and a smoothness constraint to ensure a smooth surface. Frankot and Chellappa enforced integrability to recover depth for which second partial derivatives are equal, showing higher accuracy and efficiency than Horn and Brooks method [9]. Kimmel and Sathian used a fast marching method to recover depth, resulting in a consistent, optimal, and fast algorithm for the classic SFS problem [10]. Tankus et al. re-examined SFS under perspective projection so that it could have a broader set of applications [11]. Wu et al. also considered perspective projection to solve the SFS problem, but with multiple point light sources around the camera centre [12].

The shape from focus (SFF) method can be a viable alternative single-view method for future WCE with focus controlled cameras if available in the future. SFF is a 3D reconstruction method that recovers the shape of an object using several images from one-view of the scene, captured by exploiting the focus settings of the camera. SFF uses image focus as the main cue that is measured through a focus measure (FM) operator.

The SFF method was first discussed by Nayyar et al. [13], who computed the focus value of the pixels by taking the Laplacian of the images as a first step and then used Gaussian interpolation method for shape recovery. Since then, a variety of FM operators have been proposed in the literature. These FMs are divided into six different categories (Gradient, Laplacian, Wavelet, Statistics, Discrete cosine transform and miscellaneous) based on their working principle [14]. These FMs utilize gray scale images for depth recovery. Recently, colour FMs have also been proposed that use RGB information in the images to evaluate the quality of a pixel [15], [16].

In continuation of our prior work [17], [18], both SFS and SFF methods are employed on images of an artificial colon which contains texture information. The use of artificial colon is considered due to the presence of uneven illumination, heavy lens distortion, and the continuously adapting camera response, which are common issues with WCE images. All of these factors, make it challenging to design or apply algorithms for accurate 3D reconstruction. Moreover, some 3D reconstruction algorithms cannot be tested on WCE images because of its limited design. Due to these complications, we have evaluated both SFS and SFF methods using virtual colon and camera, where we can control different parameters.

The artificial colon regions are acquired from VR-Caps [19]. VR-Caps is a virtual environment that has been designed for WCE applications. The purpose of this environment is to simulate various organ types, capsule endoscopy designs, and different tissue conditions detailed in [19]. The 3D geometry of these organs is built using computed tomography (CT) scans of patients. Whereas, the texture is created using real endoscopy images and then wrapped onto the organs. Thus, VR-Caps facilitates the testing of medical imaging algorithms for both current and future WCE designs.

VR-Caps organ models are imported into *Blender*¹ and a setup is created similar to WCE to capture images of different regions of colon with a camera and a point light source. The reconstructed shapes with both methods are then compared with ground truth data by measuring the relative root-mean-square error to estimate the accuracy of the 3D reconstruction.

WCE has LEDs which behave similar to a point light source. Therefore, images captured with such light source are relatively dark in the deeper regions due to the inverse square fall-off law. We have provided a method based on surface normals obtained from 3D reconstructed models to enhance the visibility in deeper regions of the captured images. It is also possible to change the light intensity in the image using this method. Images are enhanced by utilizing depth information recovered using both SFS and SFF methods, and then compared with each other to evaluate the accuracy of each method. Images of the same scene with different resolutions are also utilized to evaluate their impact on reconstructed models and enhanced images.

In the end, Gastroenterologists are presented with enhanced images with different light intensity, 3D models recovered using both SFS and SFF methods, and 3D models recovered with different image resolutions. 3D models and enhanced images are evaluated by gastroenterologists for quality and usefulness through subjective experiments.

This article investigates the possibility to provide enhanced viewing for gastroenterologists through 3D reconstruction and image enhancement of colon images. Single-view methods (SFS and SFF) are applied to reconstruct 3D shapes of colon images. Whereas, images are enhanced by converting

¹<https://www.blender.org/>

illumination in the image from point light to directional light using surface normals. The scope of this work is depicted in Figure 1.

The rest of the article is organized as follows. In Section II, existing SFS and SFF methods are discussed in detail that are suitable for our problem. Then we present alterations in these methods to address the problem at hand. Later, we provide a method to change the illumination in the image from point light to directional light using depth information of the scene. In Section III we discuss the ground truth models and images rendered for each method. Reconstructed surfaces from each method are also compared, along with their impact on directional light estimation. We also discuss the plausibility study for using focus controlled WCEs and evaluate the quality of 3D reconstruction for future WCE by utilizing image stacks containing different number of images. In Section IV, our discussion pertains to the subjective evaluation of both 3D reconstructed models and enhanced images. Whereas, Section V concludes the article.

II. PROBLEM FORMULATION AND METHODS

This section is divided in two parts. Initially, single-view methods that are suitable for our problem are explored. Near-light-source perspective SFS method is discussed [12], that is later applied to colon images. This method is chosen because its setup bears resemblance to that of WCE, a pill-sized device with white LEDs and an onboard camera. A colour focus measure based SFF method is chosen to apply on colon images. This method is chosen because it is proven to be more robust than other methods which utilize gray scale images for 3D reconstruction [15]. The later part of this section covers our adaptation of both of these methods to address the problem at hand.

For both SFS and SFF methods, we assume that the lens distortion is compensated for, and therefore, we have a pinhole camera model. For SFS, we assume that the light distribution is that of perfect point light and that radiometric information is obtained. To test the SFF algorithm, we assume that the focus of the camera can be controlled.

A. EXISTING METHODS

1) SHAPE FROM SHADING

SFS is a common 3D reconstruction technique that requires only one image for 3D reconstruction. SFS is convenient for 3D reconstruction of WCE images because of the known position of the light sources relative to the camera. Shape is reconstructed in two steps. At first, a reflectance model is computed based on the position of the light source, camera, and reflectance properties of the surface. In the second step, a numerical scheme is formulated using a reflectance model to solve an image irradiance equation (IIE). This numerical scheme is based on an optimization method.

A suitable method for the reconstruction task is to employ a near-light-source perspective SFS method, given the design

of the WCE device. A reflection model is computed similar to [12] which is discussed below.

Near-Light-Source Perspective SFS Method: This section describes the SFS method under point light source and perspective projection, where the light source is placed at the camera centre and the camera is pointing in the negative z -direction as shown in Fig. 2. Under perspective projection, we have,

$$x = \tilde{x} \frac{z}{f} \quad y = \tilde{y} \frac{z}{f}, \quad (1)$$

where, f is the focal length of the camera. (\tilde{x}, \tilde{y}) and (x, y, z) are image and camera coordinates, respectively.

Under the assumption of diffuse surface, radiance reflected from the diffuse surface point \mathbf{S} can be computed according to Lambert's cosine Law and inverse square fall of law of point light source [12],

$$R(\tilde{x}, \tilde{y}, z, p, q) = I_o \rho \left(\frac{\mathbf{n}(\tilde{x}, \tilde{y}, z, p, q) \cdot \mathbf{l}(\tilde{x}, \tilde{y}, z)}{r(\tilde{x}, \tilde{y}, z)^2} \right), \quad (2)$$

where, I_o is the light intensity and ρ is the surface albedo. $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ are the components of surface gradients in x and y directions, respectively. \mathbf{n} is the surface unit normal and \mathbf{l} is a unit vector representing the direction of the light ray incident at the point \mathbf{S} . $r(\tilde{x}, \tilde{y}, z)^2$ is inverse square fall-off of point light. The light source is considered at the camera centre, but can easily be extended to multiple point light source not necessarily at the camera centre [12].

The surface normal \mathbf{n} can be represented in terms of partial derivatives of the depth z with respect to x and y [7]:

$$\mathbf{n} = \frac{[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1]}{\sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1}}. \quad (3)$$

According to Horn and Brooks [7], the IIE is,

$$R(\tilde{x}, \tilde{y}, z, p, q) = I(\tilde{x}, \tilde{y}), \quad (4)$$

where, $I(\tilde{x}, \tilde{y})$ is the image irradiance. Equation (4) is solved to determine the z that minimizing the difference between image irradiance $I(\tilde{x}, \tilde{y})$ and reflectance map $R(\tilde{x}, \tilde{y}, z, p, q)$. Optimization is done over depth z whereas p and q are updated by taking the gradient of updated z . The relevant optimization problem is given by,

$$\arg \min_z E(z) = \lambda e_i(z) + (1 - \lambda) e_s(z), \quad (5)$$

where, e_i is the irradiance error and e_s represents the smoothness constraint. λ is the weighting factor between e_i and e_s and also controls the scaling as e_i and e_s are of different scale. $e_i(z)$ can be computed over the image domain $\Omega \subset \mathbb{R}^2$ as,

$$e_i(z) = \int_{\Omega} (I(\tilde{x}, \tilde{y}) - R(\tilde{x}, \tilde{y}, z, p, q))^2 d\Omega. \quad (6)$$

$e_s(z)$ is typically solved with L2 regularizer. But, we have utilized anisotropic diffusion (AD) to solve IIE. This is discussed in Section II B.

If the focus plane lies on the image plane, then $\Delta d = 0$ and image will be highly focused. The object distance (u) and image distance (v) from the lens are defined as,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \quad (7)$$

where, f is the focal length of the camera, u and v are object and image plane distances from the lens, respectively. After image acquisition, the next step is to measure the focus value of each pixel in the image stack. A colour focus measure is applied on RGB images to compute the focus value of each pixel [15].

Colour focus measure

An FM operator acts as a high-pass filter that separates the high-frequency content from low-frequency content by enhancing the focused pixels and suppressing the defocused pixels. It computes the sharpness of a pixel by selecting a local window. Object points captured with different focus settings are then compared to identify the best focused pixel for depth estimation.

To compute the focus value of pixels, a colour focus measure is applied. In the first step, the colour difference between the neighbouring pixels and centre pixels is computed and summed together in a local $\omega = 3 \times 3$ window. It is then followed by calculating their spread. The sum and the spread can be combined as [15],

$$FM_c(i, j, k) = \sigma_{\Delta}^2 \sum_{r=1}^{\omega^2-1} \delta_r, \quad (8)$$

where, δ_r is the difference between the centre and the neighbouring pixels stacked in Δ . σ_{Δ}^2 is the variance of Δ . After computing the focus value of each pixel, the depth map is obtained by finding the position of the best focused pixel which can be written as,

$$D_o(i, j) = \arg \max_k (FM_c(i, j, k)). \quad (9)$$

B. PROPOSED METHODS

This subsection contains our extension of the methods described above that is needed to cope with the relevant problem. Later, we provide a method to convert illumination in an image from point light to directional light.

1) ANISOTROPIC DIFFUSION FOR SFS

SFS is typically solved with L2 regularizer which tends to smoothen the edges because of linear diffusion. Therefore, IIE is penalized with AD [20], which is a non-linear, space-variant technique utilized to reduce the noise without smoothing edges, lines or other details which are important to interpret the resulting surface.

To impose anisotropic diffusion as a smoothness constraint, a 2×2 structure tensor is derived as a first step from

the gradient of the depth z which is given as [21],

$$S_{i,j} = \frac{\partial z}{\partial x^i} \frac{\partial z}{\partial y^j}. \quad (10)$$

Afterward, corresponding eigenvalues (λ_+ , λ_-) and eigenvectors (θ_+ , θ_-) are derived similar to [22]. From (λ_+ , λ_-) and (θ_+ , θ_-), the diffusion tensor \mathbf{D} is derived such as,

$$\mathbf{D} = \frac{\partial \psi}{\partial \lambda_+} \theta_+ \theta_+^T + \frac{\partial \psi}{\partial \lambda_-} \theta_- \theta_-^T. \quad (11)$$

In terms of (λ_+ , λ_-), Lagrangian density ψ can be written as [20],

$$e_s(z) = \int_{\Omega} \psi(\lambda_+, \lambda_-) d\Omega. \quad (12)$$

Equations (6) and (12) are combined in Equation (5) and can be written as,

$$\arg \min_z E(z) = \int_{\Omega} (\lambda(I - R)^2 + (1 - \lambda)\psi(\lambda_+, \lambda_-)) d\Omega, \quad (13)$$

The solution to Equation (13) is given by Euler-Lagrange PDE,

$$\lambda(I - R) \frac{\partial R}{\partial z} + (1 - \lambda) \nabla \cdot (\mathbf{D} \nabla z) = 0, \quad (14)$$

which we numerically solve by,

$$\frac{\partial z}{\partial t} = \nabla \cdot (\mathbf{D} \nabla z) + \frac{\lambda}{1 - \lambda} (I - R) \frac{\partial R}{\partial z}, \quad (15)$$

where, ∂t represent a small-time step, which is introduced to ensure stability with higher values of λ . The gray scale image $v(\tilde{x}, \tilde{y})$ is used to derive the structure tensor in Equation (11), that helps in simplifying and linearizing the computation of the diffusion tensor in a single step.

2) WEIGHTED L2 REGULARIZER FOR SFF

The initial depth map D_o , obtained from the focus values, contains many inaccurate depth points. This is because some areas in the colon images are smooth and therefore, the resulting images have low-frequency variations in those areas. The focus values obtained in those regions are erroneous, resulting in incorrect depth points. We correct these erroneous depth points by employing a weighted L2 regularizer.

A weighted L2 regularizer is introduced, in which the focus value of each depth point is used as a data fidelity term. Depth points containing the higher focus values are trusted and therefore retained to their actual positions. However, depth points containing smaller focus values are mistrusted, and therefore, neighbourhood depth values are given more weight to alter their position. In this way, incorrect depth points successively move closer to their original depth values.

The problem is solved by minimizing the error function E_D , which can be computed as [18],

$$E_D = \int_{\Omega} |\nabla D|^2 + \lambda FM_c |D - D_o|^2 d\Omega. \quad (16)$$

Equation (16) is solved with gradient descent such as,

$$\frac{\partial D}{\partial t} = \nabla^2 D - \lambda FM_c(D - D_o), \quad (17)$$

where λ is a weighting factor between the data fidelity term and smoothness term.

3) DIRECTIONAL LIGHT ESTIMATION

The GI tract is illuminated by white LEDs of WCE, which behave similar to point lightening that causes dim illumination in deeper regions of the captured images. The deeper regions can be illuminated if the image captured under point light is transformed to directional light.

A directional light is considered a light source at infinity, with light being directed in only one principal direction. So, there is no $(1/r^2)$ fall-off and illumination of a surface point in the image depends on surface albedo, $\rho(\tilde{x}, \tilde{y})$, and the angle between surface normals and directional light vector. Since the image captured under point light illumination follows the reflection model given in Equation (2), a point light illumination in an image can be converted to directional light I_D such as,

$$I_D = \frac{I_p r^2}{\mathbf{n} \cdot \mathbf{l}(\tilde{x}, \tilde{y}, z)} (\mathbf{n} \cdot \mathbf{L}(z)), \quad (18)$$

where, I_p is the image with point light illumination and $\mathbf{L}(z)$ is a directional light vector whose direction is $[0, 0, 1]$ assuming the light direction to be along the negative z-axis, similar to the camera. The surface normal \mathbf{n} is computed similar to Equation (3) and combined with $\mathbf{L}(z)$. The final expression for I_D can be written as,

$$I_D = I_o \rho(\tilde{x}, \tilde{y}) \frac{1}{\sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1}}, \quad (19)$$

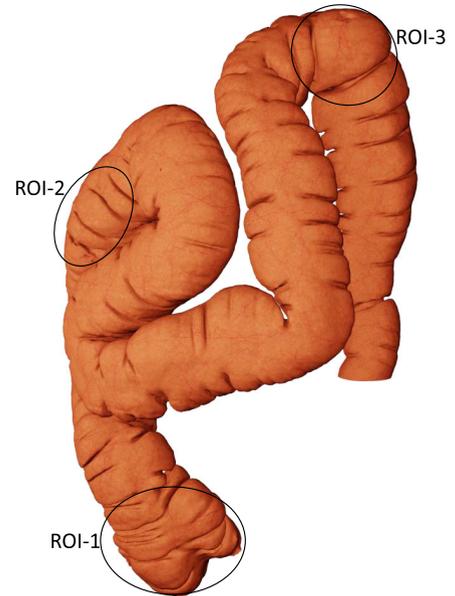
where, I_o is the light intensity which can be adjusted in the image through post-processing. This transformation method from point light to directional light can only be achieved if we possess the surface's 3D information.

III. EXPERIMENTS

A. GROUND TRUTH MODELS

Both SFS and SFF methods are tested on different areas of synthetic colon taken from VR-Caps [19]. A section of this colon is depicted in Fig. 5 (a), where the regions of interests are highlighted and shown separately in Fig. 5 (b, c, d). 3D reconstruction is usually done in medical application without the availability of ground truth data. However, we can make a ground truth scenario in *Blender* where precise comparison can be done between the reconstructed surface and the ground truth. Different parameters such as light model, focal length, focus settings of the camera etc. can also be controlled in *Blender*. This is essential to create relevant images for both methods.

To compare the reconstructed surface with the ground truth models, they are modified using the *Python API* in *Blender*. When a model is placed under a perspective camera, some of



(a) Colon model



(b) ROI-1



(c) ROI-2



(d) ROI-3

FIGURE 5. Colon model with different regions of interest.

the occluded vertices/areas are not seen by the camera. Therefore, to test the algorithm fairly, it is necessary to remove all the occluded vertices and build the model consisting of only those vertices which are inside the camera frustum and seen by the camera. The modified models are imported into *MATLAB* to compare them with reconstructed surfaces. Ground truth models are shown in Fig. 10 (a, b, c).

B. IMAGE ACQUISITION

VR-Caps models are used to acquire images for each method. Models are imported in *Blender* and then wrapped with the mucosa texture. A camera is placed on three different regions of GI. A point light source is also placed at the camera centre to imitate the illumination of WCE. A single image of each region of interest (ROI) is taken to test the SFS algorithm. Images are shown in Fig. 6

Three image stacks of the same ROIs are generated to evaluate the SFF algorithm. The only difference between the SFS and the SFF setup is a camera with adjustable focus, which is important to generate the image stack of the scene. Images are acquired by changing the focus of the camera. In each image, a certain area of the scene is kept in focus while the rest remain blurred. Different sample images of ROI-1 are shown in Fig. 7

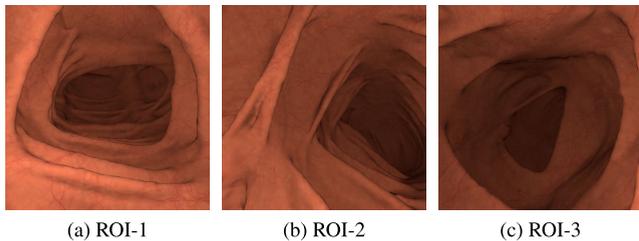


FIGURE 6. Images of all three regions for SFS.

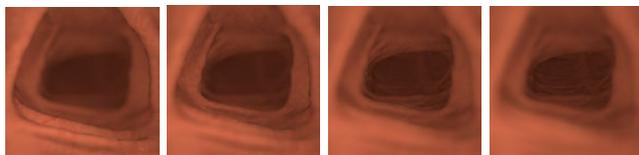


FIGURE 7. Image samples of ROI-1 with different focus for SFF.

Images are saved in Portable Network Graphic (PNG) file format for both methods. We have the possibility to control the rendered image resolution in *Blender*. Therefore, images of three different resolutions 200×200 , 500×500 , 800×800 are generated to evaluate the effect of resolution on the quality of 3D reconstruction.

While SFF only requires lens distortion correction for 3D reconstruction, SFS requires additional modifications such as texture removal and conversion of the rendered image to image irradiance to ensure correct implementation of the method. As we assume a pinhole camera model, the rendered images can be directly fed into the SFF algorithm. However, for the SFS method, we discuss the necessary modifications below.

1) IMAGE IRRADIANCE FOR SFS

Rendered images from *Blender* are converted to image irradiance $I(\tilde{x}, \tilde{y})$ to have a comparability with R computed in equation (2). This is a necessary step to correctly implement the SFS method. $I(\tilde{x}, \tilde{y})$ falling on the camera sensor is related to the gray-scale image $v(\tilde{x}, \tilde{y})$ via camera response function $r(\cdot)$ [12],

$$I(\tilde{x}, \tilde{y}) = \frac{r^{-1}[v(\tilde{x}, \tilde{y})]}{M(\tilde{x}, \tilde{y})}, \quad (20)$$

where, $M(\tilde{x}, \tilde{y})$ is the anisotropy of the light source. Point lights are perfectly isotropic by definition, and so $M(\tilde{x}, \tilde{y}) = 1$. Images are saved in Portable Network Graphics (PNG) file format and therefore, image irradiance is the gamma correction $\gamma = 2.2$ of the gray scale image [17],

$$I(\tilde{x}, \tilde{y}) = v^\gamma(\tilde{x}, \tilde{y}). \quad (21)$$

$I(\tilde{x}, \tilde{y})$ is also converted from pixel units to physical units to have corresponding units between $I(\tilde{x}, \tilde{y})$ and R . Conversion

to physical units is given by [17],

$$I_{phy}(\tilde{x}, \tilde{y}) = \frac{I(\tilde{x}, \tilde{y}) - \min I(\tilde{x}, \tilde{y})}{\max I(\tilde{x}, \tilde{y}) - \min I(\tilde{x}, \tilde{y})} \times \left(\frac{I_o \cos \theta_1}{r_1^2} - \frac{I_o \cos \theta_2}{r_2^2} \right) + \frac{I_o \cos \theta_2}{r_2^2}, \quad (22)$$

where $I_{phy}(\tilde{x}, \tilde{y})$ represents the physical value of the image irradiance and (θ_1, r_1) and (θ_2, r_2) decides the upper and lower bound of $I_{phy}(\tilde{x}, \tilde{y})$. (θ_1, θ_2) are the angles between surface normal and light ray at the maximum and minimum point on the surface, respectively. (r_1, r_2) are the distance from the light source to the maximum and minimum lit point on the surface, respectively. These points are chosen by identifying the maximum and minimum lit area in the ground truth model and then computing the angles and the distances from the light source. In general, when the ground truth models are not available, brightest and dimmest image points in physical units should be estimated for the right scale between $I(\tilde{x}, \tilde{y})$ and R . Minor deviations from the original value may not affect that much, but substantial variations can severely compromise the quality of 3D reconstruction.

2) TEXTURE REMOVAL FOR SFS

VR-Caps models are wrapped with mucosa texture [19]. The SFS method investigated in this study assumes constant albedo according to Equation (2). The SFS method might confuse the texture or veins with edges in the structure. Therefore, it is necessary to remove the texture from images as a preliminary step before employing it in the SFS algorithm.

For this purpose, Xu et al. method is utilized to separate the texture from the structure [23]. The method is based on the relative total variation (RTV) method, which captures the essential difference between texture and structure by utilizing their different properties. Later, an optimization system is developed to extract meaningful structure without prior texture information.

C. ASSESSMENT CRITERIA

The reconstructed 3D surfaces and enhanced images are assessed through objective measures and subjective tests. For objective evaluation, the reconstructed surfaces are compared with ground truth models by measuring relative root-mean-square error (r_{RMSE}). This method has been chosen to compute the depth errors with respect to a reference depth, in order to ensure that the resulting errors are easily interpretable. By employing this method, an overall geometric deformation in the 3D reconstructed model can be evaluated. An error value of 0 represents a perfect 3D reconstruction, while a value of 1 indicates a highly distorted 3D reconstruction. r_{RMSE} can be computed as,

$$r_{RMSE} = \frac{1}{d_{max}} \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{D}_i - D_i|^2}, \quad (23)$$

where, D is the ground truth depth, d_{max} is the max ground truth depth point and \hat{D} is the depth of the recovered 3D shape.

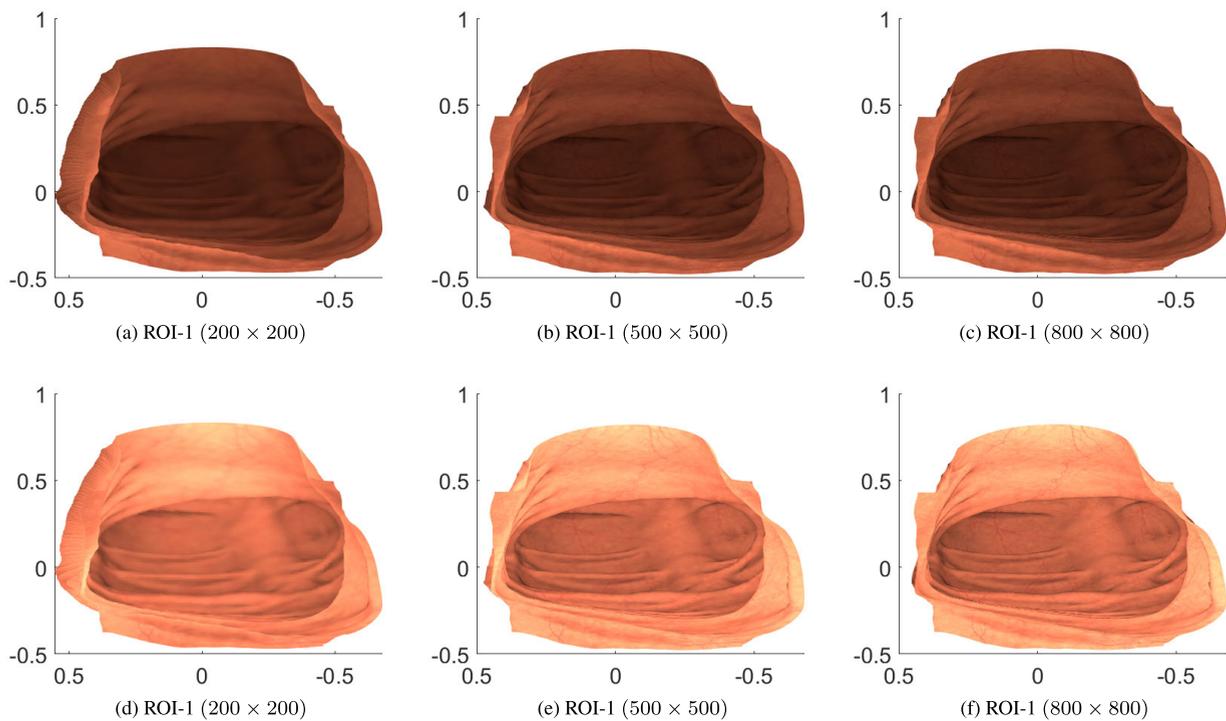


FIGURE 8. Recovered models with SFS: Texture-wrapped illumination with point light and directional light.

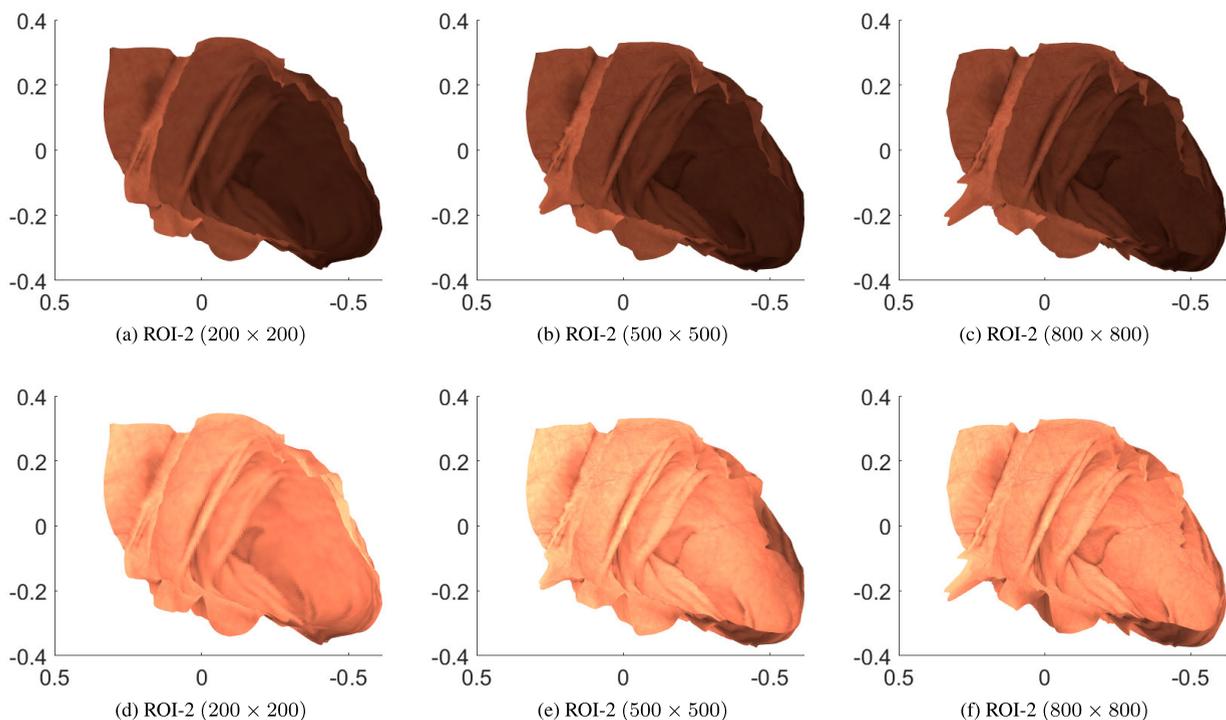


FIGURE 9. Recovered models with SFF: Texture-wrapped illumination with point light and directional light.

n represents the total number of depth points considered for error estimation.

For subjective evaluation of 3D reconstructed models and enhanced images, three experiments have been tested with five gastroenterologists. The experiments are discussed in Section IV in detail.

D. RESULTS

After acquiring the image stack for SFF and single image of each region for SFS, they are fed to their respective algorithms for 3D reconstruction. For the SFS method, images are first gamma corrected using Equation (21) and then converted to physical units using Equation (22). Texture is then removed

TABLE 1. Relative RMSE of reconstructed models for both SFS and SFF methods.

Colon Regions	Resolutions	r_{RMSE} (SFS)	r_{RMSE} (SFF)
ROI-1	200 × 200	0.1793	0.1627
	500 × 500	0.1789	0.1469
	800 × 800	0.1773	0.1451
ROI-2	200 × 200	0.1368	0.0871
	500 × 500	0.1328	0.0698
	800 × 800	0.1316	0.0641
ROI-3	200 × 200	0.1719	0.1571
	500 × 500	0.1714	0.1397
	800 × 800	0.1641	0.1296

from the images using Xu et al. method. An initial reflectance map is computed from Equation (2) using a flat surface as an initial depth z . Depth values z are updated using Equation (15) where, p and q are updated as a gradient of z such as, $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$. The value of λ is different for different cases and is empirical in our experiment. The SFS models recovered with different image resolutions of ROI-1 wrapped with corresponding original point-lit texture are shown in Fig. 8 (a, b, c).

For the SFF method, the colour image stack of each region is utilized to compute the focus value of each pixel using equation (8). An initial depth map D_o is reconstructed by finding the position of the best focused pixel in the image stack. D_o along with its focus values, is then utilized in Equation (17) to correct for inaccurate depth points. The value of λ is different for different cases and empirical in our experiment. Finally, the (x, y) coordinates are computed using Equation (1). The SFF models recovered with different resolution of ROI-2 wrapped with original point-lit texture are shown in Fig. 9 (a, b, c).

Both SFS and SFF methods are compared with each other by computing r_{RMSE} between reconstructed surfaces and ground truth models. Table 1 shows that the SFF method achieves lower r_{RMSE} for all three regions. r_{RMSE} also decreases with increasing resolution for both methods, as seen in Table 1.

Both SFS and SFF methods have successfully been able to recover the 3D shape of the colon images. However, it can be clearly seen from Fig. 10 and Table 1 that the accuracy of the SFF method is better than the SFS. The SFS method does not seem to handle the deeper portion in all three regions and smoothen them all, whereas the SFF method retains the level of detail in the reconstructed regions consistently from top to bottom.

After recovering the shapes with both methods, images of the colon are enhanced to provide better contrast in the deeper regions. Originally, models were illuminated with a point light source, which causes dim illumination in the deeper regions of the captured images. By converting the illumination from point light to directional light, contrast is enhanced significantly.

Depth recovered with both SFF and SFS methods are applied to change the illumination of the images. Surface

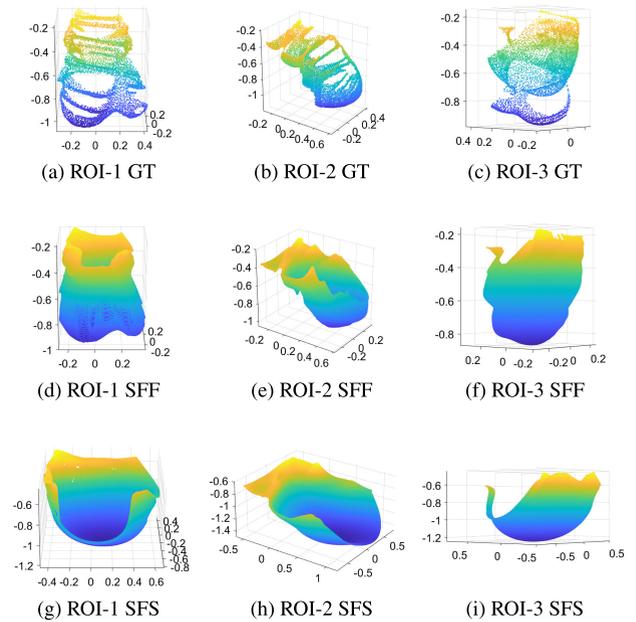


FIGURE 10. Ground truth and recovered models with SFF and SFS.

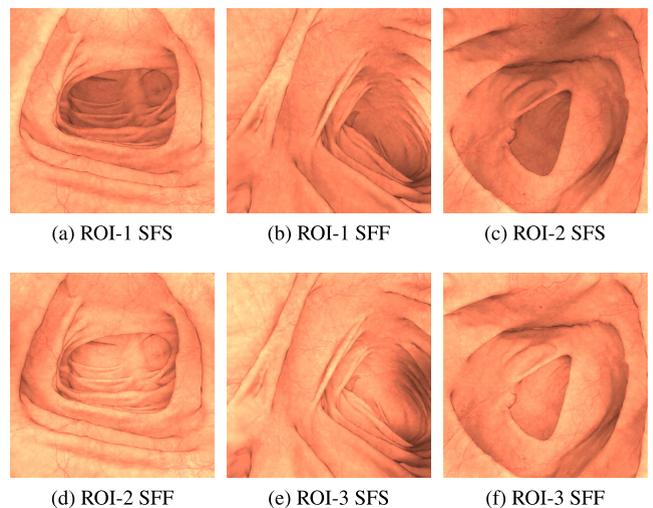


FIGURE 11. Contrast enhanced images.

normals (\mathbf{n}), point light vector (\mathbf{l}), and attenuation effect of point light source ($1/r^2$) are computed using depths recovered from each method. These values are used in Equation (19) to convert the illumination in the original image from point light to directional light. The results are shown in Fig. 11. Enhanced images attained with depth information of the SFS method are comparatively darker in deeper regions than the SFF. This is because the SFS method is unable to reconstruct the deeper regions accurately. Recovered models wrapped with enhanced images for different resolution of ROI-1 and ROI-2 are shown in Fig. 8 (d, e, f) and Fig. 9 (d, e, f) for SFS and SFF respectively.

Image resolution has a significant effect not only on the quality of 3D reconstruction, but also in computing the

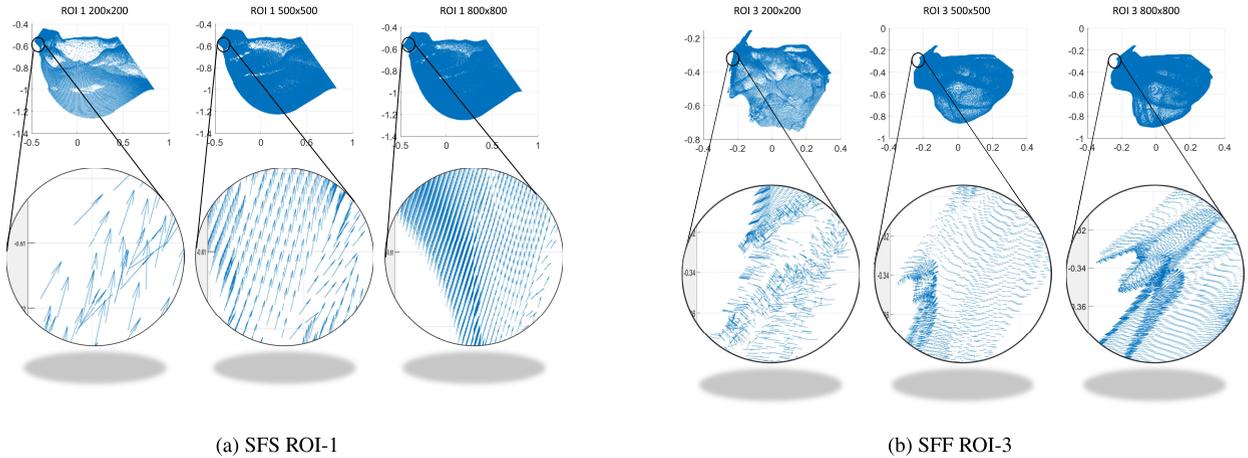


FIGURE 12. Normal orientations of reconstructed surfaces for different resolutions.

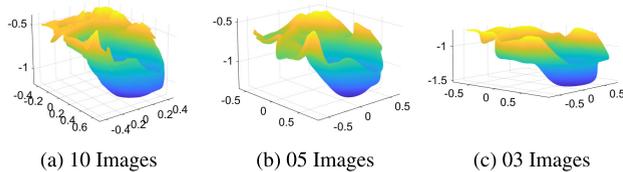


FIGURE 13. Recovered models with SFF with varying number of images.

directional light illumination. The effect does not look significant from Table 1 as the error measures is an average over the entire model, effectively suppressing small differences in detail between different resolutions. Surface normals are cluttered in some areas in the reconstruction, as shown in Fig. 12. The clutter effect is reduced as the resolution increases. The effect of resolution can also be seen by noticing the bottom shape of ROI-3 in Fig. 12 (b) in particular. The effect is not significant in Fig. 12 (a) as the SFS method smoothens the deeper regions of the reconstructed surfaces. The effect of resolution can also be noticed by comparing the texture lit with directional light in Fig. 9 (d, e, f). The deeper regions in Fig. 9 (d) are darker compared to the upper regions. The light gets more and more uniform as the resolution increases to 500×500 and 800×800 .

E. PLAUSIBILITY STUDY FOR FOCUS CONTROLLED WCE

We have seen from the above results that the SFF method provides better 3D reconstruction than the SFS. However, SFF requires cameras that can adjust their focus settings to capture images at different focal planes. This can be achieved by either changing the distance between the camera and the object or adjusting the camera’s focus settings.

At present, the WCEs that are available for use in GI imaging do not have cameras with focus control settings. These WCE’s rely on the natural contractions of the GI tract to move through the digestive system. To address this limitation,

TABLE 2. Relative RMSE of reconstructed models with varying numbers of images for SFF.

Colon Region	No. of Images	r _{RMSE}
ROI-2	03	0.1399
	05	0.0983
	10	0.0919

various prototypes have been proposed in the literature for controlling the movement of WCE. Glass et al. developed a mechanism to anchor the WCE to the intestinal walls [24]. Karagozler et al. also presented a six-legged WCE designed to mimic a crawling motion [25]. These advancements represent promising steps towards the development of capsules that can be precisely controlled and maneuvered for more effective imaging and diagnosis.

Technological advancements have made it feasible to incorporate a mechanical system within WCE to control the focus of the lens by adjusting the distance between the lens and sensor. Such a mechanism is already employed in modern mobile phones, which have narrow thickness. Therefore, it is plausible to introduce this technology to WCE to adjust the focus settings of the camera.

To perform depth recovery using SFF, it is necessary to obtain an image stack containing numerous images with a constant step size, Δ_{step} , to ensure accurate depth estimation. However, capturing such many images with varying focus settings can be power-intensive and create battery problems in WCE. By carefully adjusting the step size, it is possible to obtain an image stack that enables accurate depth recovery while minimizing the impact on the WCE’s battery life. To optimize the image capturing process, the step size can be varied as depth of field (DOF) has a direct relation with object distance (u) for a given f-number (N), as follows [26],

$$DOF = \frac{2u^2Nc}{f^2}. \tag{24}$$

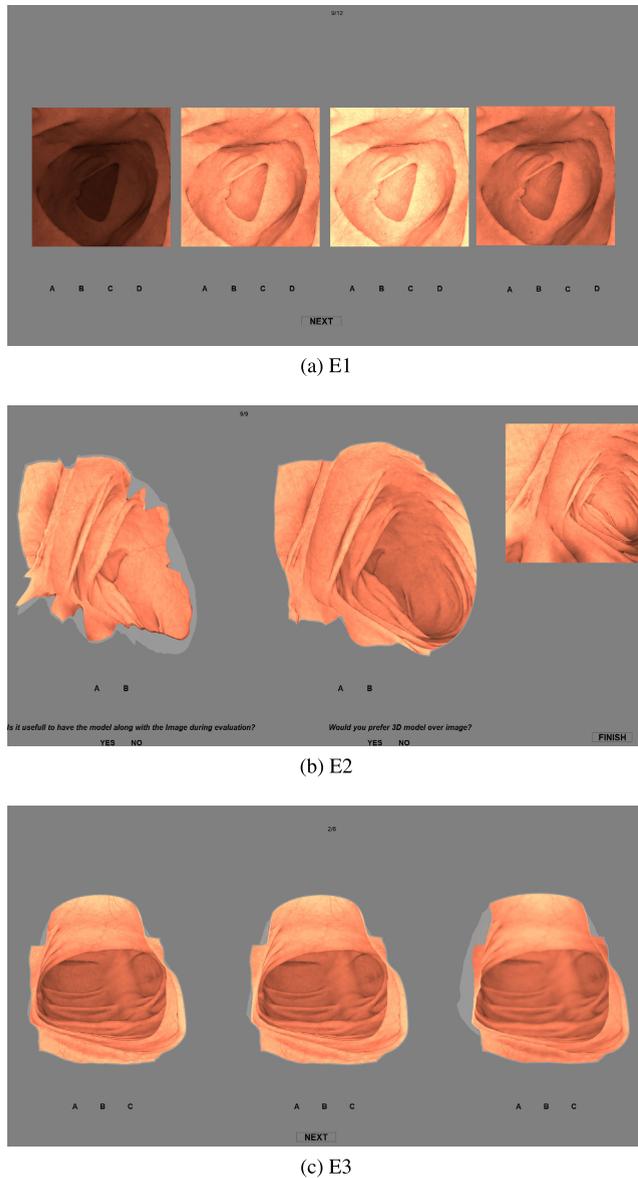


FIGURE 14. Experiments for subjective evaluation.

According to equation (24), DOF is proportional to u^2 . Consequently, capturing images with varying step sizes Δ_{step} would be the most optimal solution for both power consumption and SFF algorithm.

We test the plausibility of focus controlled WCE by utilizing different sets of image stacks containing different number of images. Since it is very difficult to capture a numerous image stack, we tested our hypothesis on ROI-2 by recovering depth with three different image stacks containing 10, 05, 03 images. The results are shown in Fig. 13. With ten images, a 3D shape is reconstructed that closely approximates the ground truth. The reduction in the number of images results in a gradual decline in the quality of 3D reconstruction. Table 2 shows the r_{RMSE} of ROI-2 when compared with ground truth. The results show that r_{RMSE} increases as the number of images decreases. Using only three images results in a

lower quality reconstruction compared to the SFS whereas, with five images, the quality is much better than the SFS. This confirms that future technological advancements may enable the widespread application of the SFF method for 3D reconstruction of human GI regions captured using WCE.

IV. SUBJECTIVE EVALUATION

The quality and usefulness of the enhanced images and 3D reconstructed models with both SFS and SFF methods are evaluated by gastroenterologists. In total, three experiments were designed,

- 1) **E1**: The original images of the colon are compared with three contrast enhanced images, lit with directional light of three different strengths (low ($DL-N$), medium ($DL-M$), High ($DL-H$)), where, the strength of each enhanced image is adjusted such as, $DL-N = 1.25I_o$, $DL-M = 1.75I_o$, $DL-H = 2.25I_o$. An example is shown in Fig. 14 (a).
- 2) **E2**: The reconstructed models with the SFS and the SFF methods are compared for different resolutions. An example is shown in Fig. 14 (b). Questions are also asked about the usefulness and preference of the 3D model over its image.
- 3) **E3**: The reconstructed models of the same region, recovered using different image resolutions, are compared with each other for both methods. An example is shown in Fig. 14 (c).

The experiments were carried out with one gastroenterologist in a pilot study. Based on the results of this trial, light strength in enhanced images is adjusted. Some new questions are also added. Updated experiments are then conducted with five gastroenterologists at “Innlandet Hospital Trust Gjøvik” (SI Gjøvik).

A. DESCRIPTION of EXPERIMENT

1) APPLICATION AND SETUP

The subjective experiments are created using *MATLAB GUIDE* [27]. In **E1**, four images (original and enhanced with different strength of directional light) are placed horizontally side by side in random order for each trial. Twelve trials are done in total, with images of 200×200 and 500×500 resolutions for all three ROIs. In **E2**, reconstructed surfaces with both SFS and SFF are placed side by side in random order for each trial. An enhanced image of the corresponding region is also placed in the top-right corner. Nine trials are done in total for all three ROIs. In **E3**, reconstructed surfaces of a single region with three different resolutions are placed side by side in random order for each trial. Six trials are done in total for both methods. Both in **E2** and **E3**, reconstructed models are wrapped with *light gray* color (i.e., RGB values of [153, 153, 153]) on the outside to avoid confusion. The gastroenterologists have the possibility to *rotate* and *zoom* the models and view them from different angles.

3D models and images are displayed on a *Dell ultra-sharp 27"* monitor (U2719DC) over a *middle gray* (i.e., RGB values of [119, 119, 119]) background. The experiments are

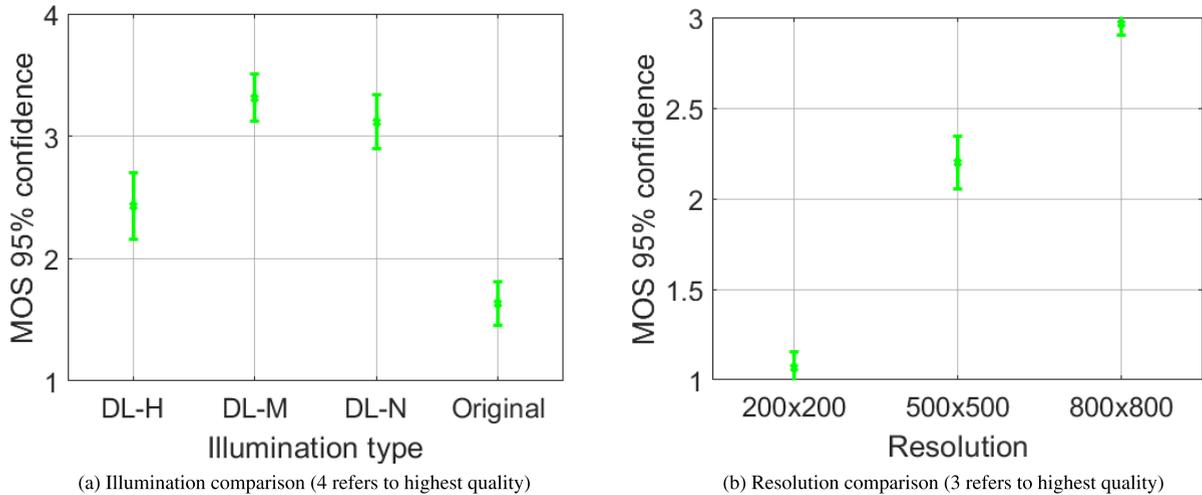


FIGURE 15. MOS score for E1 and E3.

conducted in a room at SI Gjøvik with the same type of lighting condition as the room used for assessment of colonoscopy images, that is D65 lighting. Therefore, the monitor is calibrated for D65 lighting. The experiments are time-limited due to the tight schedule of gastroenterologists. Therefore, a limited but representative set of enhanced images and 3D models are chosen for subjective evaluation.

2) ASSESSMENT

Each candidate is asked to make the following assessments for the different experiments.

- 1) **E1:** The candidates are asked to categorize the images from *A* to *D*, with *A* being the highest and *D* being the lowest quality, corresponds to 4 to 1 points on a linear scale. Several images can be given the same score.
- 2) **E2:** The candidates are given a binary choice to rate reconstructed models by giving scores *A* and *B*, with score *A* being more preferred over score *B*. Candidates are also asked about the usefulness and preference of the reconstructed 3D models in two separate questions. The decision for usefulness is “yes” if a candidate finds it useful to have the 3D model along with the image during evaluation. Otherwise, the candidate should click “no”. The decision for preference is “yes” if a candidate prefers the 3D model over the original image. Otherwise, the candidate should click “no”.
- 3) **E3:** The candidates are asked to categorize the reconstructed models from *A* to *C*, with *A* being the highest and *C* being the lowest quality, corresponds to 3 to 1 points on a linear scale. Several models can be given the same score.

3) INFORMATION ON CANDIDATES

The five gastroenterologists have long experience, but somewhat different background, which is as follows:

- **Candidate 1:** Expert in assessment of WCE images and has some experience in evaluating 3D models.

- **Candidate 2:** Some experience in assessment of WCE images, but no experience in evaluating 3D models.
- **Candidate 3:** Some experience in assessment of WCE images, but no experience in evaluating 3D models.
- **Candidate 4:** Some experience in assessment of WCE images, but no experience in evaluating 3D models.
- **Candidate 5:** No experience in assessment of WCE images or 3D models.

B. RESULTS

Mean opinion score (MOS) is computed for E1 and E3 as shown in Fig. 15. The MOS is computed as the arithmetic mean of all scores corresponding to the grades of each experiment. Our analysis involves computing a 95% confidence interval assuming that the distribution of variation in the mean follows a normal distribution [28]. Similarly, Wilson interval (WI) with 95% confidence interval is computed for E2 as shown in Fig. 16. Results of the pilot study are not included in computing the score, as it was conducted mainly to improve the experiment.

In E1, gastroenterologists clearly preferred enhanced images to original images where DL-M achieves the best score, as shown in Fig. 15 (a). Often, DL-H is preferred over the original, despite the former being intensely bright. In E2, gastroenterologists preferred the SFS results over the SFF as shown in Fig. 16 (a) even though it has less accuracy when compared with ground truth models. In E3, models recovered with 800 × 800 images are preferred over others, as shown in Fig. 15 (b). Candidates also find it very useful to have a 3D model along with the image and in more than 40% of the cases they prefer the 3D model over its image as shown in Fig. 16 (b).

1) DISCUSSION ON RESULTS

Overall, the enhanced images are preferred over the original in E1, while DL-M is marginally preferred over DL-N. It is worth mentioning that one of the candidates reported the usefulness of varying levels of light intensity depending on

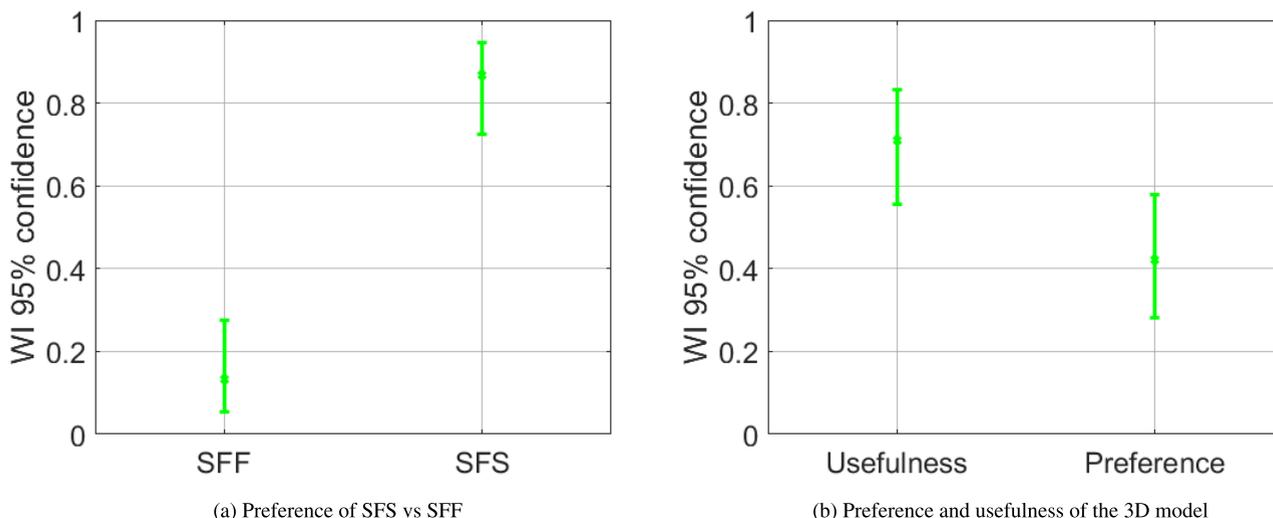


FIGURE 16. WI score for E2.

the type of disease in the image, such as strong lighting for inflammation and medium/low lighting for bleeding. This presents an advantage for our method, as we can control the light strength through post-processing.

Surprisingly, models recovered with the SFS are mainly preferred and considered more convenient than the SFF, despite the former being less accurate when compared with ground truth models. This is likely because the models recovered with SFF contain more detail and less smoothness, resulting in significant stretching of some pixels during the texture-to-structure warping process. While the texture appearance is smooth due to the smoother structure of models recovered with the SFS method, which may make them more perpetually pleasing.

Models recovered with 800×800 images are given the best score because of the quality of the structure and clarity of the mucosa in the texture. Models recovered with 200×200 images are given the least score due to noise in the structure and blurry texture. Models recovered with 500×500 images are clearly preferred over, 200×200 and in some cases they are considered of the same quality as models recovered with 800×800 images.

It is interesting that the candidates find it very useful to include the 3D models during evaluation and in more than 40% of the cases preferred 3D models over its image. It is also worth mentioning that one of the candidates expressed interest in only evaluating the 3D models reconstructed from infected colon images. Therefore, the candidate has given zero score to both “usefulness” and “preference” questions. Despite this, we achieved good results indicating the importance of 3D models in medical application.

In single-view methods, an image does not provide information about occluded regions. As a result, when constructing the 3D shape, the depth information in those regions can be deceptive. This is a concern raised by one of the gastroenterologists while evaluating the 3D models. One way of addressing this problem is to combine multi-view methods

with single-view methods, where depth in occluded areas can be corrected using the information in the successive images, where those regions may be visible. But this is only possible if those areas are captured by WCE. When information on occluded regions is unavailable, it may be helpful to shade regions of missing data in the 3D model with a distinct colour to avoid misguidance or confusion.

2) COMMENT ON THE NUMBER OF CANDIDATES

Ideally, many candidates should have performed the subjective evaluation of 3D models and enhanced images. However, it is difficult to obtain results from a sufficient number of gastroenterologists within a narrow time frame due to limited availability of the qualified personnel. Since the results are consistent in most cases after the five candidates we managed to contact, we chose to conclude the experiment.

V. CONCLUSION AND FUTURE WORKS

We have investigated the possibility of reconstructing colon images with single-view methods (SFS and SFF). Image data sets are generated using a virtual environment consisting of synthetic human gastrointestinal (GI) regions. Results show that both methods can handle the complexity of GI images, but the SFF method preserves detail better than the SFS. Therefore, the SFF method will be convenient for 3D reconstruction of WCE images if focus controlled WCEs are available in the future.

We have also discussed a method based on surface normals obtained from a 3D model to better illuminate and thereby enhance contrast in captures of larger depths by changing the illumination from point light to directional light. Results show that the method can improve lighting quite well in deeper regions of the captured images. Since the method depends on depth information, its accuracy depends upon how accurately depth is recovered from its image(s). The SFF method can illuminate the deeper regions better than the SFS

because the SFF method provides more accurate 3D models, as shown in the results.

Through experiments, we have shown that the resolution has a significant effect on the quality of both 3D reconstructed models and contrast enhanced images. The accuracy of the 3D models and uniformity of the light in enhanced images gets better as the resolution increases. We have also provided a plausibility study for the potential use of the SFF method in future WCE design with focus controlled cameras. Our study suggests that the SFF can yield promising results and encourages further investigation in this context.

We have also conducted a comprehensive investigation with gastroenterologists to assess the efficacy of 3D models in clinical evaluations. The results show that the use of 3D reconstructed models, in addition to images, are found to be highly beneficial, and in some cases, the gastroenterologists even preferred 3D models over the original images. Our study also shows the usefulness of light conversion method as the strength of the lighting can be controlled, which is very useful when evaluating images containing different disorders.

In future work, different techniques can be evaluated to achieve improved texture warping onto the surface. Alternatively, it is possible to stitch together textures from different images to minimize pixel stretching. Single-view methods can be combined with multi-view methods to correct for depth information in areas that are not visible in the original image due to occlusion. The SFS method will also be applied on real WCE images where we will have to deal with different textures, specularities, and distorted images.

ACKNOWLEDGMENT

The authors would like to thank Prof. Per Farup for his valuable suggestions to improve the subjective experiments. They would also like to thank M.D. Simen Svendsen Vatn, M.D. Alexander Ernst Endres, M.D. and Ph.D. Snorri Olafsson, and M.D. Heidi Kvarme for participating in the subjective experiments.

REFERENCES

- [1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, p. 417, May 2000.
- [2] H. Ham, J. Wesley, and H. Hendra, "Computer vision based 3D reconstruction: A review," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 9, no. 4, p. 2394, Aug. 2019.
- [3] T. Collins and A. Bartoli, "3D reconstruction in laparoscopy with close-range photometric stereo," in *Proc. MICCAI*, 2012, pp. 634–642.
- [4] P. A. Floor, I. Farup, and M. Pedersen, "3D reconstruction of the human colon from capsule endoscopy video," in *Proc. Colour Vis. Comput. Symp. (CVCS)*, Gjøvik, Norway, 2022.
- [5] A. Koulaouzidis, D. Iakovidis, D. Yung, E. Mazomenos, F. Bianchi, A. Karagyris, G. Dimas, D. Stoyanov, H. Thorlacius, E. Toth, and G. Ciuti, "Novel experimental and software methods for image reconstruction and localization in capsule endoscopy," *Endoscopy Int. Open*, vol. 6, no. 2, pp. E205–E210, Feb. 2018.
- [6] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsafaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, K. Huang, K. S. Nikita, B. P. Veasey, M. Zervakis, J. H. Saltz, and C. S. Pattichis, "AI in medical imaging informatics: Current challenges and future directions," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020.
- [7] B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," *Comput. Vis., Graph., Image Process.*, vol. 33, no. 2, pp. 174–208, Feb. 1986.
- [8] K. Ikeuchi and B. K. P. Horn, "Numerical shape from shading and occluding boundaries," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 141–184, Aug. 1981.
- [9] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 439–451, Jul. 1988.
- [10] R. Kimmel and J. A. Sethian, "Optimal algorithm for shape from shading and path planning," *J. Math. Imag. Vis.*, vol. 14, no. 3, pp. 237–244, May 2001.
- [11] A. Tankus, N. Sochen, and Y. Yeshurun, "Shape-from-shading under perspective projection," *Int. J. Comput. Vis.*, vol. 63, no. 1, pp. 21–43, Jun. 2005.
- [12] C. Wu, S. G. Narasimhan, and B. Jaramaz, "A multi-image shape-from-shading framework for near-lighting perspective endoscopes," *Int. J. Comput. Vis.*, vol. 86, nos. 2–3, pp. 211–228, Jan. 2010.
- [13] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.
- [14] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, May 2013.
- [15] H. Mutahira, B. Ahmad, M. S. Muhammad, and D. R. Shin, "Focus measurement in color space for shape from focus systems," *IEEE Access*, vol. 9, pp. 103291–103310, 2021.
- [16] M. T. Mahmood and U. Ali, "Measuring focus quality in vector valued images for shape from focus," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 3399–3405.
- [17] B. Ahmad, P. A. Floor, I. Farup, and M. Kresović, "3D reconstruction of gastrointestinal regions from single images," in *Proc. Colour Vis. Comput. Symp. (CVCS)*, Gjøvik, Norway, 2022, pp. 1–11.
- [18] B. Ahmad, I. Farup, and P. A. Floor, "3D reconstruction of gastrointestinal regions using shape-from-focus," in *Proc. 15th Int. Conf. Mach. Vis. (ICMV)*, Jun. 2023, pp. 463–470.
- [19] K. Incecan, I. O. Celik, A. Obeid, G. I. Gokceler, K. B. Ozyoruk, Y. Almalioglu, R. J. Chen, F. Mahmood, H. Gilbert, N. J. Durr, and M. Turan, "VR-caps: A virtual environment for capsule endoscopy," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101990.
- [20] D. Tschumperle and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 506–517, Apr. 2005.
- [21] S. Di Zeno, "A note on the gradient of a multi-image," *Comput. Vis., Graph., Image Process.*, vol. 33, no. 1, pp. 116–125, Jan. 1986.
- [22] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Trans. Image Process.*, vol. 5, no. 11, pp. 1582–1586, 1996.
- [23] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–10, Nov. 2012.
- [24] P. Glass, E. Cheung, and M. Sitti, "A legged anchoring mechanism for capsule endoscopes using micropatterned adhesives," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 12, pp. 2759–2767, Dec. 2008.
- [25] M. E. Karagozler, E. Cheung, J. Kwon, and M. Sitti, "Miniature endoscopic capsule robot using biomimetic micro-patterned adhesives," in *Proc. 1st IEEE/RAS-EMBS Int. Conf. Biomed. Robot. Biomechtron. BioRob.*, Jan. 2006, pp. 105–111.
- [26] E. Allen and S. Triantaphillidou, *The Manual of Photography*. Boca Raton, FL, USA: CRC Press, 2012.
- [27] *Create Apps With Graphical User Interfaces in MATLAB*. Accessed: Feb. 15, 2023. [Online]. Available: <https://www.mathworks.com/discovery/MATLAB-gui.html>
- [28] D. R. Bull, *Measuring and Managing Picture Quality*. New York, NY, USA: Academic, 2014.



BILAL AHMAD received the B.S. degree in engineering sciences from the GIK Institute of Science and Technology, Pakistan, in 2014, and the M.S. degree in electrical and computer engineering from Sungkyunkwan University, South Korea, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Norwegian University of Science and Technology (NTNU), Norway. His research interests include 3D shape recovery and image processing.



PÅL ANDERS FLOOR received the B.Sc. degree in electrical engineering from Gjøvik University College, in 2001, and the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), in 2003 and 2008, respectively. He held several postdoctoral positions with the Intervention Center, Oslo University Hospital, the Institute of Clinical Medicine, the University of Oslo,

as well as the Department of Electronics and Telecommunications, NTNU, from 2008 to 2015. Since 2016, he has been with the Colourlab, Department of Computer Science, NTNU, lately as a Researcher. His research interests include information theory and signal- and image processing with applications within the medical field.



ØISTEIN HOVDE received the Doctor of Medicine (M.D.) degree from the University of Oslo, in 1984. He is currently an Associate Professor with the Institute of Clinical Medicine, University of Oslo, and a Professor with the Norwegian University of Science and Technology (NTNU). He is also a Senior Consultant (Gastroenterologist) with Innlandet Hospital Trust, Gjøvik, Norway. He is also the Head of the Gastroenterological Division, Innlandet Hospital

Trust. His Ph.D. focused on epidemiology in inflammatory bowel diseases (IBD: Crohn's disease and ulcerative colitis). His clinical work is focused on inflammatory bowel diseases, endoscopy, and therapeutic endoscopy. For many years, he has had different positions in national and European gastrointestinal organizations, such as the Norwegian Gastroenterological Association and the United European Gastroenterological Organization (UEG).

•••



IVAR FARUP received the M.Sc. degree in technical physics from the Norwegian Institute of Technology, in 1994, and the Ph.D. degree in applied mathematics from the University of Oslo, Norway, in 2000. He is currently a Professor of computer science with the Norwegian University of Science and Technology (NTNU), with a focus on color science and image processing.